STATISTICS
TEXAS A&M UNIVERSITY

Course Introduction and Background on Surveys and Time Domain Astronomy

September 14, 2016

Course Information

What is Astrostatistics?

Time Domain Astronomy Background

Surveys

Outline

Course Information

What is Astrostatistics?

Time Domain Astronomy Background

Surveys

Tentative Schedule

Date	Topic
Sept 7	Introduction, Time Domain Astronomy (James Long)
Sept 14	Time Domain Astronomy (James Long)
Sept 21	Time Domain Astronomy (James Long)
Sept 28	Bootstrap, Hypothesis Testing (Jogesh Babu)
Oct 5	Bootstrap, Hypothesis Testing (Jogesh Babu)
Oct 12	Spectral Data, Bayesian SED Fitting (Viviana Acquaviva)
Oct 19	Something Bayesian (Tom Loredo)
Oct 26	Approximate Bayesian Computation (Jessi Cisewski)
Nov 2	Spectral Data, Approximate Models (James Long)
Nov 9	Spectral Data, Approximate Models (James Long)
Nov 16	Student Presentations
Nov 23	NO CLASS
Nov 30	Student Presentations

Where to Find Course Information

- course website: `http://stat.tamu.edu/~jlong/astrostat`
- email me: `jlong` + at + `stat.tamu.edu`
- office hours: after class or by appointment
- course lectures: videos on samsi website

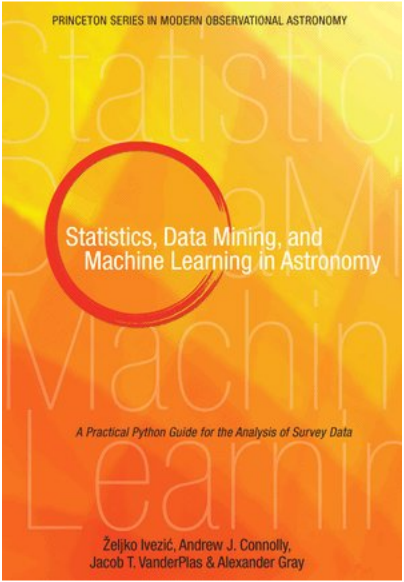
Grading / Work Load

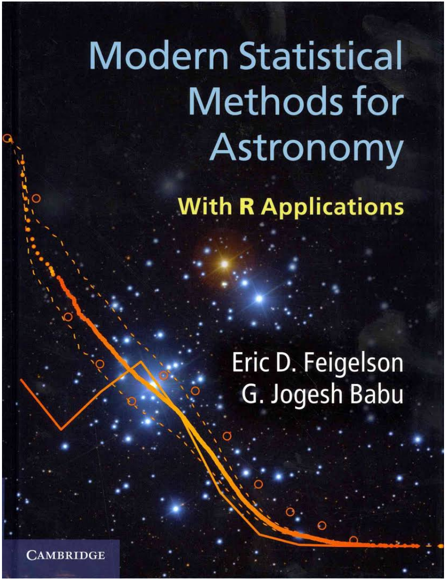
- P/F for everyone
- sign up sheet in back of classroom
- no homework
- suggested readings on course website
- target audience: 1st year statistics PhD student
- course project
  - choose topic by mid October, submit 1 paragraph description
  - meet with me in early November to discuss progress
  - $\approx$ 30 min in class presentations on Nov 16,30
  - more details at later lectures

Project Topic Suggestions

- reproduce article results / apply to new data set
  - "Unsupervised Transient Light Curve Analysis Via Hierarchical Bayesian Inference" by Sanders, ApJ
  - "Modeling lightcurves for improved classification . . ." by Faraway, Statistical Analysis and Data Mining
  - "Some Aspects of Measurement Error . . . " Kelly ApJ
  - "A flexible method of estimating luminosity functions" Kelly ApJ
- tutorial on research / coding tools
  - bayesian computing, eg stan
  - ipython / jupyter notebooks
- simulation study, algorithm comparison
- overview of topic in astronomy we did not cover**
  - photometric redshift estimation
  - source extraction from images
  - statistics of gravity waves

Astrostatistics Textbooks





Outline

Course Information

What is Astrostatistics?

Time Domain Astronomy Background

Surveys

Common Themes

- complex, multi–stage inference problems
  - “A framework for statistical inference in astrophysics” [4]
- measurement error, heteroskedastic errors with known variance
- large data sets
- data types: images, spectra, maps
- complex scientific questions that require close astronomer–statistician collaboration

Statistical Methods Used: Pretty Much Everything

- machine learning
- measurement error models
- samplers, MCMC
- functional data analysis
- time series
- hierarchical models
- spatial statistics

Getting Involved in Astrostat Research @ SAMSI

WG 1: Uncertainty Quantification and Astrophysical Emulation

Leaders: Derek Bingham (SFU) and Earl Lawrence (LANL)
To join email: wg1@sakai.duke.edu

WG 2: Synoptic Time Domain Surveys

Leaders: Ashish Mahabal (Caltech) and G. Jogesh Babu (PSU)
To join email: wg2@sakai.duke.edu
Weekly meeting time: Tuesdays at 12:00-1:00PM, EDT

WG 3: Multivariate and Irregularly Sampled Time Series (MISTS)

Leaders: Ben Farr (U.Chicago) and Soumen Lahiri (NCSU)
To join email: wg3@sakai.duke.edu

WG 4: Astrophysical Populations (AP)

Leaders: Jessi Cisewski (Yale) and Eric Ford (PSU)
To join email: wg4@sakai.duke.edu
Weekly meeting time: Thursdays at 12:00-1:00PM, EDT

WG 5: Statistics, Computation, and Modeling in Cosmology (COSMO)

WG Leaders: Jeff Jewell (JPL) and Joe Guinness (NCSU)
To join email: wg5@sakai.duke.edu

Outline

Course Information

What is Astrostatistics?

Time Domain Astronomy Background

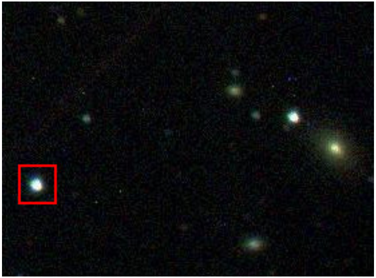
Surveys

Telescopes Take Images of the Night Sky

Images to Catalogs

For every object in every image, need to identify and characterize objects:

- location (RA/Dec)
- brightness (magnitudes)
- filter (or band)
- time (MJD)

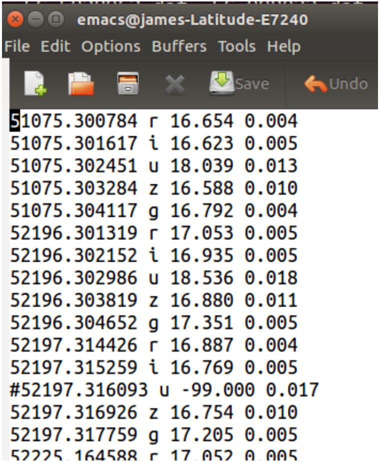


Hard task (possible project):

- "SExtractor: Software for source extraction" Aanda, Bertin [1]
- "The SDSS Imaging Pipeline" Lupton [2]
- "Celeste: Variational inference for a generative model of astronomical images" ICML, Regier [3]

Time Domain: Objects are Observed Over Time

When telescope takes repeated images of the same area of the sky, we obtain many brightness measurements for each object.




```
51075.300784 r 16.654 0.004
51075.301617 i 16.623 0.005
51075.302451 u 18.039 0.013
51075.303284 z 16.588 0.010
51075.304117 g 16.792 0.004
52196.301319 r 17.053 0.005
52196.302152 i 16.935 0.005
52196.302986 u 18.536 0.018
52196.303819 z 16.880 0.011
52196.304652 g 17.351 0.005
52197.314426 r 16.887 0.004
52197.315259 i 16.769 0.005
#52197.316093 u -99.000 0.017
52197.316926 z 16.754 0.010
52197.317759 g 17.205 0.005
52225.164588 r 17.052 0.005
```

Outline

Course Information

What is Astrostatistics?

Time Domain Astronomy Background

Surveys

Surveys

Surveys:

- Non–survey: Hey, there's something interesting over there, let's point our telescope at it.
- Survey: Observations are scheduled in advance with some science goals in mind.

Tradeoffs in Surveys:

- narrow vs wide
- shallow vs deep
- number of filters
- temporal coverage

Given a certain budget of telescope time, these are optimized for one (or several) science goals.

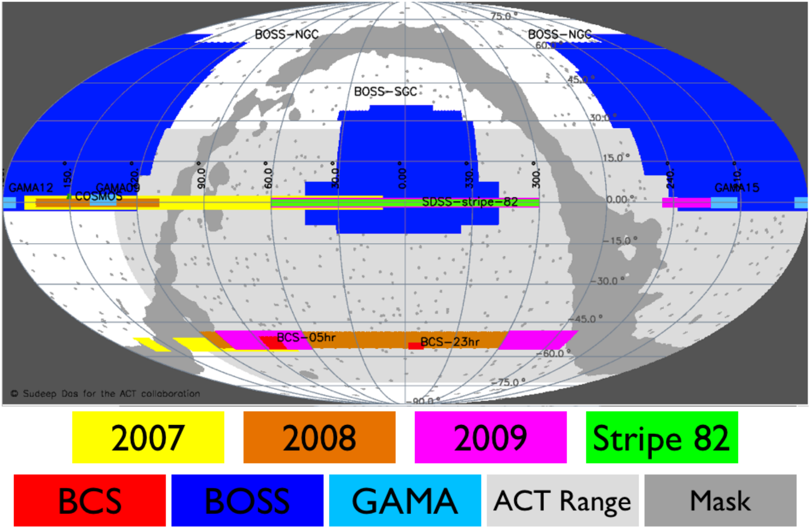
Tom Loredo's Talk on Surveys:

`https://www.samsi.info/wp-content/uploads/2016/08/Tutorial-CosmicDemographics_Loredo.pdf`

Major Time Domain Optical Surveys

- Sloan Digital Sky Survey – Stripe 82 (SDSS)
- Optical Gravitational Lensing Experiment (OGLE)
- Kepler
- Dark Energy Survey (DES)
- Large Synoptic Survey Telescope (LSST)

SDSS – Stripe 82

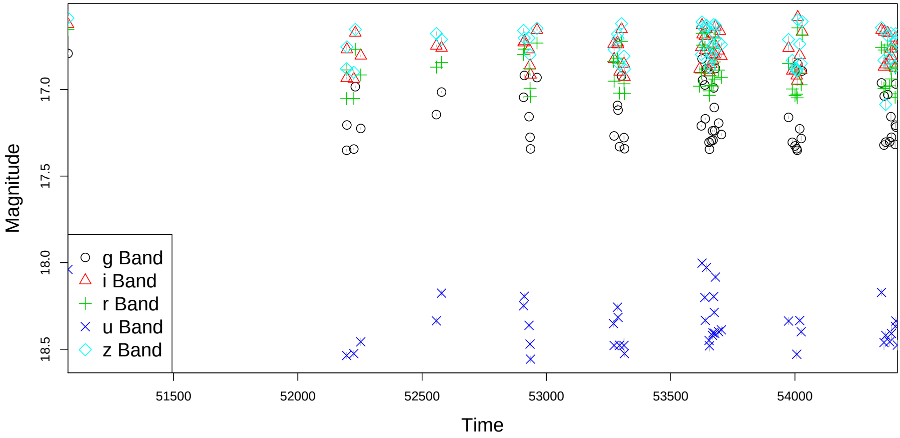


See Sesar [6] and Sesar [5]
Image credit: `https://inspirehep.net/record/859503/files/F3.png`

SDSS Photometry

- standard star catalog ($\approx$ 1 million): http://www.astro.washington.edu/users/ivezic/sdss/catalogs/stripe82.html
- variable source catalog ($\approx$ 67,000): http://www.astro.washington.edu/users/ivezic/sdss/catalogs/S82variables.html

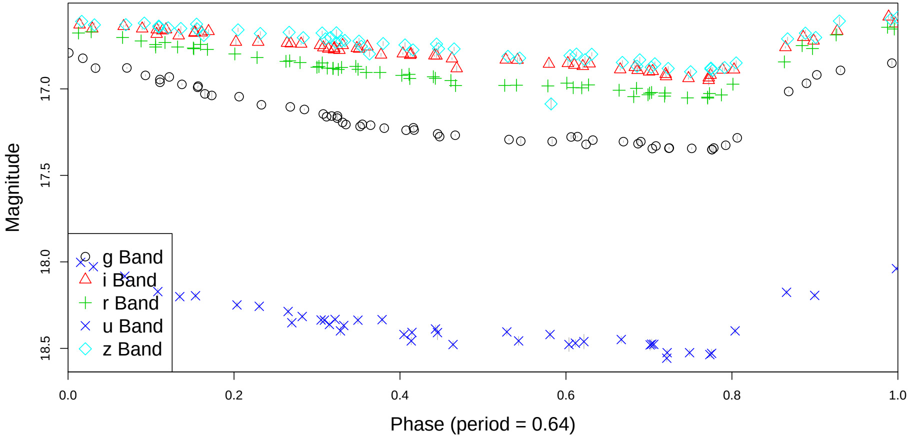
SDSS Variable Light Curve



- ≈ 67,000 of these in SDSS Stripe 82
- ≈ 70 observations / filter in 5 (u,g,r,i,z) filters
- ≈ 10 years

Folded Light Curve

Estimate period of 0.64 days, plot magnitude versus phase for object.



RR Lyrae variable star, useful for determining distance.

Selecting Variable Sources

- Data for single star is

$$D = \{(t_{jb}, m_{jb}, \sigma_{jb})\}_{j=1}^{n_b}$$

for $b = 1, \ldots, B$. Observe brightness m_{jb} at time t_{jb} with uncertainty σ_{jb} in filter (band) b.
- If a star is constant then a possible model is

$$m_{jb} = \mu_b + \epsilon_{jb}$$

where $\epsilon_{jb} \sim N(0, \sigma_{jb}^2)$.
- $\widehat{\mu}_b = (\sum \sigma_{jb}^{-2})^{-1} \sum_{j=1}^{n_b} \frac{m_{jb}}{\sigma_{jb}^2}$

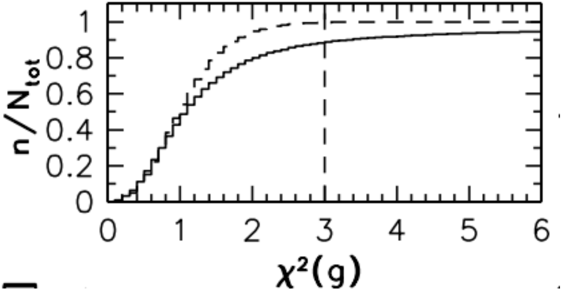
$$RSS(b) = \sum_{i=1}^{n_b} \left(\frac{m_{jb} - \widehat{\mu}_b}{\sigma_{jb}} \right)^2$$

- Large RSS(b) suggest variable.

Using RSS to Make Cuts

For example Sesar 2007 [6] defines

$$\chi^2(g) = \frac{1}{n_g - 1} RSS(g)$$



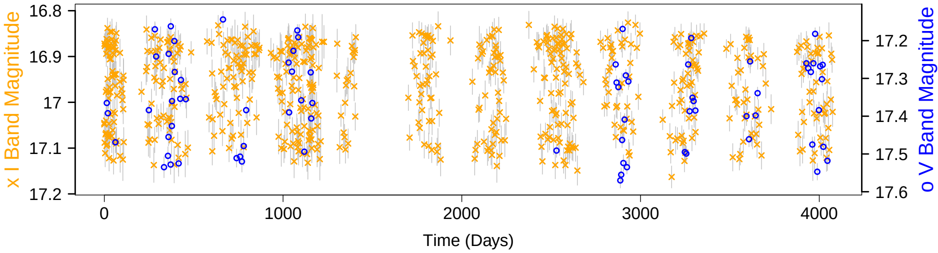
A source must have $\chi^2(g)$ above 3 to be labeled variable.

Optical Gravitational Lensing Experiment (OGLE)

Surveyed Small Magellanic Cloud, Large Magellanic Cloud, and Galactic Bulge.

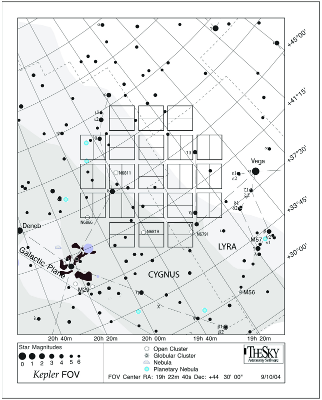
OGLE Photometry

- OGLE–III catalog of variables
  `http://ogledb.astrouw.edu.pl/~ogle/CVS/`
- ≈ 400,000 variable stars
- 2 filters (I,V)
- more data in I



Goal: Find transiting extrasolar planets.

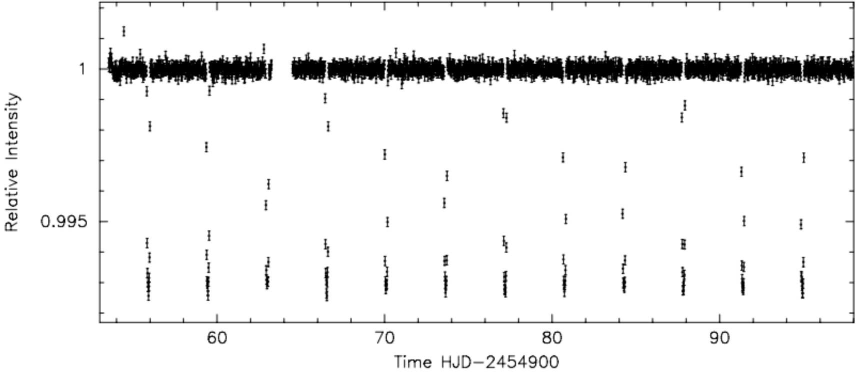
- observation every 30 minutes
- one filter
- $\approx 145,000$ main sequence stars



Transiting Method for Detecting Exoplanets:
`http://kepler.nasa.gov/multimedia/Interactives/HowKeplerDiscoversPlanetsElementary/flash.cfm`

Credit for Images: `https://en.wikipedia.org/wiki/Kepler_(spacecraft)`

Kepler Photometry



Planet transits every ≈ 3.5 days.

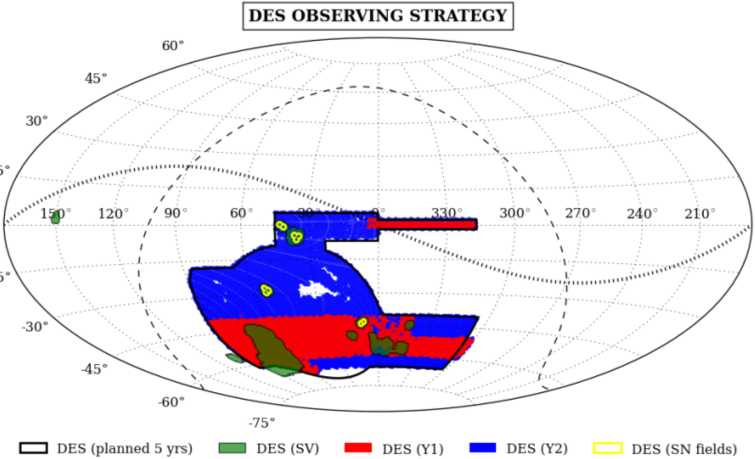
Dark Energy Survey

Goal: Investigate Large Scale Structure of Universe using

- Type Ia supernovae
- Baryon Acoustic Oscillations (BAO)
- Number of galaxy clusters
- Weak gravitational lensing

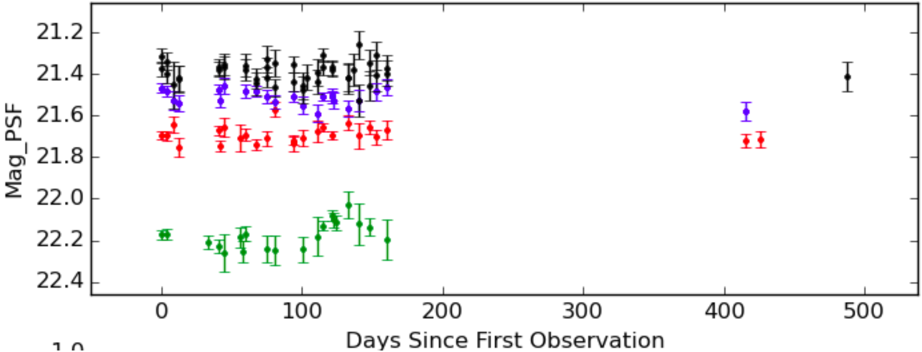
What is a good observing strategy for supernovae?

Dark Energy Survey Sky Coverage



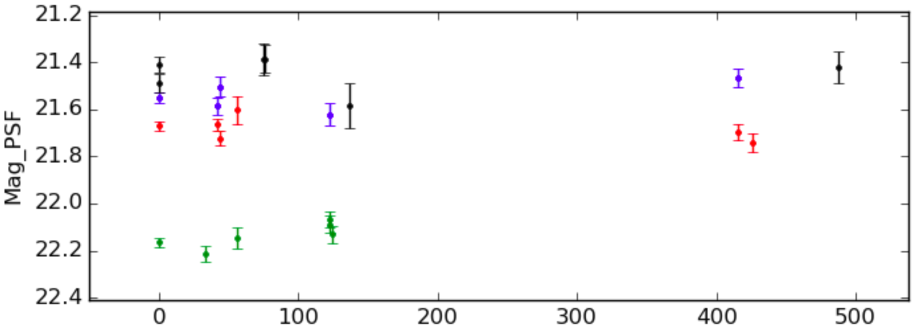
Taken from: `https://www.darkenergysurvey.org/the-des-project/survey-and-operations/`

Dark Energy Survey – Supernovae Fields Y1



Many observations / object but fewer objects.

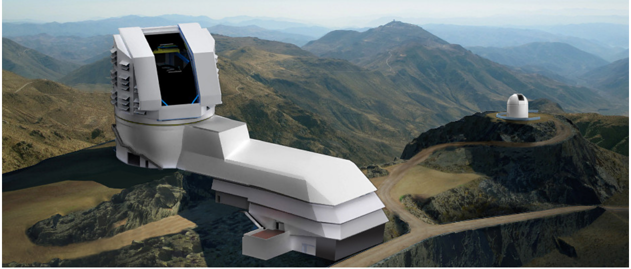
Dark Energy Survey – Wide Fields Y1



Few observations / object but many objects (≈ 100 million stars).

Large Synoptic Survey Telescope

- first light in 2019
- wide range of science goals
- unprecedented quantity of data



Summary

- large (and growing) amount of data
- different survey strategies to target different science goals
- data types: two modes of thinking
  - (vector valued) irregularly sampled **time series**
  - (vector valued) irregularly sampled **functional data**
- heteroskedastic magnitude error, "known" variance

Bibliography I

[1] Emmanuel Bertin and Stephane Arnouts. Sextractor: Software for source extraction. *Astronomy and Astrophysics Supplement Series*, 117(2):393–404, 1996.

[2] Robert Lupton, James E Gunn, Zeljko Ivezic, Gillian R Knapp, Stephen Kent, and Naoki Yasuda. The sdss imaging pipelines. *arXiv preprint astro-ph/0101420*, 2001.

[3] Jeffrey Regier, Andrew Miller, Jon McAuliffe, Ryan Adams, Matt Hoffman, Dustin Lang, David Schlegel, and Mr Prabhat. Celeste: Variational inference for a generative model of astronomical images. In *Proceedings of the 32nd International Conference on Machine Learning*, 2015.

[4] Chad M Schafer. A framework for statistical inference in astrophysics. *Annual Review of Statistics and Its Application*, 2:141–162, 2015.

[5] Branimir Sesar, Željko Ivezić, Skyler H Grammer, Dylan P Morgan, Andrew C Becker, Mario Jurić, Nathan De Lee, James Annis, Timothy C Beers, Xiaohui Fan, et al. Light curve templates and galactic distribution of rr lyrae stars from sloan digital sky survey stripe 82. *The Astrophysical Journal*, 708(1):717, 2010.

[6] Branimir Sesar, Željko Ivezić, Robert H Lupton, Mario Jurić, James E Gunn, Gillian R Knapp, Nathan De Lee, J Allyn Smith, Gajus Miknaitis, Huan Lin, et al. Exploring the variable sky with the sloan digital sky survey. *The Astronomical Journal*, 134(6):2236, 2007.